# Research Based on sample size calculation and Monte Carlo simulation

Yao Wang *

Faculty of Information, Liaoning University, Shenyang, China, 110031

* Corresponding Author Email: 15649522005@139.com

**Abstract.** This paper focuses on the inspection of parts and components and the quality control of finished products in the production process of electronic products, and focuses on how to reduce the number of inspections as much as possible to ensure that the defective rate of parts and components meets the requirements of enterprises and how to find the optimal production strategy in different situations. We used the sample size calculation formula and Monte Carlo simulation method to determine the minimum sample size under the reliability of 95% and 90% respectively, and simulated the detection process under the condition of different defective rates for enterprises to make reasonable decisions according to different situations. In addition, we designed the optimal detection and decision strategy for different production stages. The research results provide a basis for enterprises to formulate reasonable quality inspection and control programs, and have significant practical value.

**Keywords:** Sample size calculation, Monte Carlo simulation, Optimal detection.

## 1. Introduction

In modern manufacturing industry, product quality control is one of the core links to ensure the market competitiveness and brand reputation of enterprises. Especially in the production process of electronic products, the quality of spare parts directly determines the qualification rate and overall performance of finished products. Due to the high cost and cumbersome operation of quality inspection, enterprises are facing the challenge of how to reduce the cost of inspection while ensuring the quality of finished products. Therefore, establishing effective sampling inspection methods, reasonably evaluating the defective rate of parts and components, and formulating treatment strategies for unqualified products are the key issues to improve production efficiency and control costs [1]. This demand leads to the study of sample size calculation and quality control decision in production stage.

In the research of quality control and sampling inspection, scholars have proposed a variety of optimization methods. For example, for the optimization of sampling inspection scheme, some studies accurately calculate the minimum sample size required by statistical models to ensure that the defective rate can be effectively evaluated at different reliability levels [2]. In addition, the Monte Carlo simulation method can evaluate the uncertainty under complex production conditions, so as to optimize the sampling strategy and detection process [3]. As for multi-stage production decision-making, previous studies have discussed how to establish an optimized quality control mechanism between parts inspection and finished product assembly, and proposed a phased inspection model [4]. At the same time, some studies have analyzed the cost-effectiveness of the disassembly and reuse strategy, and proposed the decision-making basis for the disassembly and re inspection of unqualified products after the inspection of finished products [5]. In addition, some literatures have studied the cumulative effect of defective rate in the detection link, and proposed optimization strategies for multi link and multi process [6].

Although the existing research has made significant progress in quality control and sampling inspection, there are still some limitations. For example, many studies mainly focus on a single detection link, less systematic consideration of the multi-stage decision-making problem in the whole production process, so it is difficult to provide a comprehensive optimization scheme. Moreover, most of the existing researches on the disassembly and inspection strategies of unqualified products are static analysis, which fail to fully consider the changes of cost and benefit in the dynamic production environment [7].Our research overcomes the limitations of existing research, introduces the multi-stage decision-making framework of quality control and sampling inspection, and dynamically considers the changes of costs and benefits in the production environment. This method makes the disassembly and inspection strategy more comprehensive and more adaptive, and more in line with the actual needs.(The original data required are all from https://www.mcm.edu.cn/html_cn/node/a0c1fb5c31d43551f08cd8ad16870444.html.)

## 2. Description of model principle

### 2.1. Composition of sample size calculation formula

The sample size calculation formula is a vital component in statistical analysis, as it determines the minimum number of samples required to accurately estimate population parameters, such as proportions or means, within a specified confidence level and margin of error. Its accurate application is essential in various fields of research, including experimental design, surveys, and statistical inference. This formula not only ensures the precision and reliability of the research findings but also helps avoid the waste of resources by preventing the use of unnecessarily large sample sizes. Proper sample size determination is thus fundamental for the success and efficiency of research projects.

In the context of proportion estimation, the sample size calculation typically relies on the normal approximation to the binomial distribution, a widely accepted method in statistical practice. This process involves several key steps:

(1) **Determine the confidence level**: The researcher selects an appropriate confidence level, such as 95%, which corresponds to a critical value from the standard normal distribution, denoted as Z. This value defines the range within which the true population parameter is expected to lie with the chosen level of confidence.

(2) **Estimate the population proportion**: If prior data or information is available, it can be used to estimate the population proportion, denoted by p. If no prior information is available, the most conservative estimate of p=0.5 is used to ensure the largest possible sample size, thus safeguarding the study's power.

(3) **Set the allowable error**: The researcher must also define the allowable margin of error (denoted as E), which represents the maximum acceptable difference between the sample estimate and the true population value. The choice of E depends on the research objectives and desired precision.

(4) **Calculate the sample size**: Once the confidence level, population proportion estimate, and allowable error are determined, these values are substituted into the sample size calculation formula. For proportion estimation, the sample size n is calculated using the formula. This equation ensures that the sample size is sufficient to estimate the population parameter within the defined margin of error, with the specified confidence level.

(5) Interpret the results: The computed sample size represents the minimum number of observations required to make statistically reliable estimates of the population proportion. Using a sample size smaller than this would risk underestimating the true population parameters, while using a larger sample size may unnecessarily increase resource consumption without substantial gains in accuracy.

We define a formula for calculating the confidence interval of the overall proportion.

$$C_i = p \pm Z_x \times \sqrt{\frac{p(1-p)}{n}}, \tag{1}$$

where $C_i$ is confidence interval, $p$ represents frequency of occurrence of a feature or event in the sample, $Z_x$ means the $Z$ value of the standard normal distribution corresponding to the given set signal level, and $n$ indicates the number of samples.

We also define the cumulative distribution function of binomial distribution, which is used to calculate the cumulative probability that the value of random variable $x$ is greater than or equal to $k$.

$$P(x \geq k) = \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i}, \tag{2}$$

where $x$ is the number of successful trials in $n$ independent trials, $k$ refers to the cumulative probability that the number of successful tests does not exceed $k$ in $n$ tests, $n$ represents the total number of independent Bernoulli tests, $i$ indicates all possible values of success times, which varies from 0 to $k$, and $p$ means the probability of success of each independent test.

We also define a sample size calculation formula for estimating the proportion.

$$n = \frac{Z^2 \times p(1-p)}{E^2}, \tag{3}$$

where $n$ is the sample size, $Z$ represents the critical value of the standard normal distribution corresponds to the selected confidence level, $p$ means the estimated proportion of the population with certain characteristics, and $E$ indicates the half width of confidence interval.

The sample size calculation formula is essential in research design and statistical inference, ensuring accurate and reliable results while optimizing resource use. Selecting an appropriate sample size is critical for valid conclusions, preventing underpowered studies with weak findings and avoiding unnecessary resource expenditure with excessively large samples. By determining the sample size based on a predefined confidence level and margin of error, researchers can generalize findings to the broader population, ensuring statistically significant and meaningful conclusions, which are crucial for informed decision-making in fields such as public health, social sciences, and market research. Ultimately, the sample size calculation formula plays a pivotal role in balancing the demands of precision, reliability, and resource efficiency in scientific research and practical applications [8].

## 2.2. Overview of Monte Carlo simulation

Monte Carlo simulation is a computational technique that uses random sampling and statistical analysis to model systems characterized by uncertainty and variability. By generating random variables, it approximates the effects of uncertainty and estimates a range of possible outcomes, providing valuable insights into risk. This technique is particularly useful in scenarios where traditional methods are insufficient to capture system complexities. Monte Carlo simulation is widely applied in fields like physics, finance, engineering, and operations research, allowing researchers to assess system behavior probabilistically and make informed decisions. It provides deeper insights into variability and risk, making it indispensable for analyzing complex problems involving uncertainty. The first step in Monte Carlo simulation is to establish a clear mathematical or computational model that accurately represents the system, identifying key output variables for analysis.

(1) **Define input distributions**: The next step is to define probability distributions for the key input variables. These distributions reflect the uncertainty and variability in the system and are based on historical data, theoretical assumptions, or expert judgment. Correctly defining these distributions is crucial for ensuring that the simulation accurately captures potential scenarios.

(2) **Random sampling**: A random number generator is then used to sample from the input distributions repeatedly, simulating a range of possible conditions. Each set of sampled inputs represents a unique scenario, and this process is repeated many times, allowing the simulation to explore a wide array of outcomes.

(3) **Calculate results**: For each set of sampled inputs, the model computes the corresponding output values. These results are recorded for every iteration, building a dataset of simulated outcomes

that reflect the system's behavior under various conditions.

(4) **Statistical analysis**: Finally, statistical analysis is conducted on the simulation results. Key metrics such as expected values, variances, and confidence intervals are calculated to provide insights into the system's overall behavior and the risks involved. This analysis helps quantify the likely outcomes and supports informed decision-making.

Monte Carlo simulation's strength lies in its ability to handle complex problems with high levels of uncertainty or variability. Unlike deterministic models that provide a single outcome, Monte Carlo simulation produces a distribution of possible outcomes, offering a probabilistic perspective that is more reflective of real-world conditions. The method is particularly useful for risk analysis, allowing researchers and decision-makers to evaluate the probability and magnitude of different risks. The accuracy of Monte Carlo simulation is closely related to the number of iterations performed—the larger the number of simulations, the more reliable the results. However, it is important to balance computational resources with the required precision, as increasing the number of iterations can be computationally expensive. Monte Carlo simulation is widely used in various fields due to its flexibility and robust analytical capabilities. In finance, it is employed to model the uncertainty in market prices, assess portfolio risks, and value financial derivatives. In engineering, it aids in reliability analysis and risk assessment of complex systems. Operations research leverages Monte Carlo simulation to optimize supply chains, production processes, and scheduling under uncertainty. The method's applicability across different domains highlights its indispensable role in modern decision-making and problem-solving, especially in scenarios where uncertainty plays a significant role in determining outcomes.

We define the formula for calculating the estimated expected value.

$$E\left|f(x)\right| \approx \frac{1}{N}\sum_{i=1}^{N} f(x_i), \tag{4}$$

where $N$ is the total number of simulations, $x_i$ indicates the sample extracted from the distribution of $X$, $X$ means a variable with a certain probability distribution, and $f(x_i)$ refers to the value of function $f$ at $x_i$.

We also define the integral estimation formula of Monte Carlo simulation.

$$I \approx \frac{1}{N}\sum_{i=1}^{N} f(x_i), \tag{5}$$

where $I$ is the integral value we want to estimate, $N$ means the total number of sampling points, and $f(x_i)$ indicates a point randomly selected from the integral region.

We still define the formula for calculating Monte Carlo standard deviation.

$$SE = \frac{\sigma}{\sqrt{N}}, \tag{6}$$

where $SE$ is the standard error, $\sigma$ represents the estimated value of sample variance and the dispersion degree of sample results, and $N$ refers to the number of samples.

The Monte Carlo simulation method is fundamentally based on repeated random sampling to approximate the true solution of a problem. By simulating a large number of scenarios, it enables the quantification of variability and uncertainty in system behavior [9]. This approach is particularly valuable for analyzing complex systems, where traditional analytical methods may struggle to handle uncertainty or intricate dynamics. Through the application of relevant mathematical formulas, Monte Carlo simulation estimates key statistical measures such as expected values, variances, and confidence intervals. These estimates provide insight into the likely outcomes and the degree of uncertainty associated with them. As a result, Monte Carlo simulation is a highly flexible and powerful analytical tool, especially in contexts where uncertainty plays a significant role, making it an indispensable method for solving problems that are otherwise difficult to model or predict accurately [10].

## 3. Results

### 3.1. Analysis of sample size and simulation test

In this section, we calculate the minimum sample size under 95% and 90% reliability by using the sample size calculation formula, and simulate the sampling process under different defective rate scenarios. The results are shown in Table.1 and Figure.1.

**Table 1.** Minimum sample size corresponding to different reliability

| Reliability | Minimum sample size |
|---|---|
| 95% | 865 |
| 90% | 609 |

From the table, we can draw the following conclusion: on the premise of 10% of the nominal value, if the defective rate of parts and components is determined to exceed the nominal value under 95% reliability, the batch of parts and components will be rejected, and the corresponding minimum sample size is 865; If it is determined that the defective rate of spare parts exceeds the nominal value under 90% reliability, the batch of spare parts will be rejected, and the corresponding minimum sample size is 609.

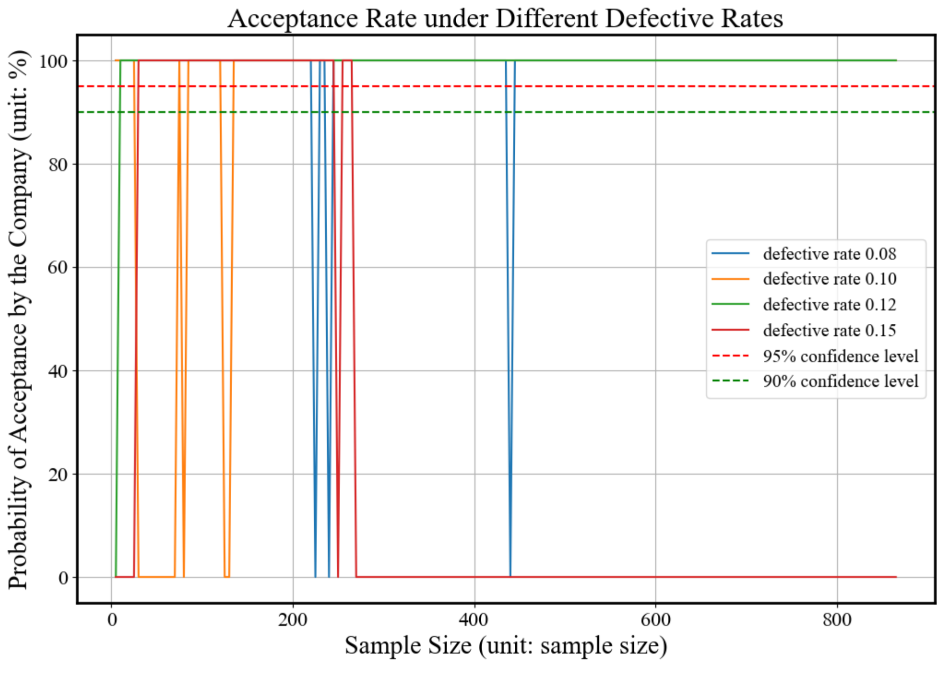


**Figure 1.** Acceptance Rate under Different Defective Rates

The figure shows four curves corresponding to four different defective rates (0.08, 0.10, 0.12, 0.15). It can be observed that with the increase of sample size, the acceptance probability under different defective rates gradually tends to be stable, which is consistent with the theory in statistics that the increase of sample size will improve the estimation accuracy.

(1) For the case that the defective rate is 0.08, the enterprise has been able to accept this batch of spare parts with a high probability under a small sample size (about 5 to 50 samples). With the further increase of sample size, the acceptance probability increases rapidly, and is close to 1. This indicates that when the actual defective rate is lower than the nominal defective rate (0.10), the enterprise has

a high probability of accepting the batch of spare parts.

(2) For the case where the defective rate is 0.10, that is, the actual defective rate is equal to the nominal defective rate, and the acceptance probability gradually approaches 90% with the sample size. This means that under the condition that the nominal defective rate is 10%, the enterprise can still accept the batch of products with high confidence under a large sample size.

(3) When the defective rate rises to 0.12 and 0.15, the acceptance probability decreases significantly with the increase of sample size, especially when the sample size is large (more than 200 pieces), the enterprise has a high probability of rejecting this batch of parts. This is because the actual defective rate is significantly higher than the nominal defective rate, and the sampling test results show that this batch of products do not meet the quality requirements.

The figure also contains two dashed lines, corresponding to the confidence levels of 95% and 90%, respectively, as the reference standard for enterprises to accept spare parts.

(1) The red dotted line indicates the 95% confidence level, which means that under the 95% confidence level, the enterprise has a high probability to reject the parts with the actual defective rate exceeding the nominal value. The role of this line is to help enterprises decide whether to reject products under strict quality control. According to the results in the figure, with the increase of sample size, when the actual defective rate is higher than the nominal defective rate (such as 0.12 and 0.15), the acceptance probability of the enterprise is gradually lower than the 95% confidence level, indicating that the enterprise should reject this batch of accessories.

(2) The green dotted line indicates the 90% confidence level, which means that under the 90% confidence level, the enterprise has a high probability to accept the parts with the actual defective rate close to or lower than the nominal value. The dotted line indicates that when the actual defect rate is close to the nominal defect rate (such as 0.10), the enterprise has a high probability of accepting this batch of spare parts, especially in a large sample size.

### 3.2. Key analysis of decision-making scheme

In order to quantify the impact of each decision on the total cost of production, we constructed a mathematical model to represent each component of the total cost, and combined with Monte Carlo simulation to deal with the uncertainty existing in reality (such as the fluctuation of defective rate). The results are shown in the Figure.2.

According to the chart, the following is a detailed analysis of the expected total cost, combined with the selection of the optimal strategy. We simulated all possible strategy combinations and found the optimal strategy for each case.

(1) **Case 1 to Case 3**: In these scenarios, the optimal strategy is strategy 3, which involves detecting and disassembling both parts 1 and 2, as well as the finished products. This strategy minimizes the total cost by effectively identifying defects at both the component and product levels, thus preventing defective products from reaching the market.

(2) **Case 4**: The optimal strategy in this case is strategy 1, where there is no need to test either spare part 1 or 2, nor disassemble them. Instead, only the finished products are tested. This approach minimizes costs by reducing unnecessary inspections of components when the testing of the final product is sufficient to maintain quality control.

(3) **Case 5**: Here, the optimal strategy is strategy 11, which focuses on testing only spare part 2, while skipping the testing and disassembly of finished products. This strategy effectively reduces costs by targeting the component most likely to have defects, thus ensuring quality without incurring the costs associated with full-scale disassembly.

(4) **Case 6**: In this scenario, strategy 16 is identified as optimal. This strategy completely forgoes all detection and disassembly operations, relying solely on the defective rate of the parts. In this case, skipping all processing steps minimizes costs, indicating that additional testing or disassembly would not provide sufficient cost-benefit given the low risk of defects.

(5) **Defective rate and strategy adjustment**: Strategy 3, selected for cases 1 to 3, highlights that when both spare parts and finished products exhibit high defective rates, comprehensive detection

and disassembly significantly reduce the exchange losses caused by defective products. In contrast, cases 4, 5, and 6 demonstrate that when the defective rate is low or the cost of disassembly is high, the optimal strategies tend to involve minimizing or skipping detection and disassembly. This highlights the dynamic relationship between defective rates and the chosen detection and disassembly strategies.

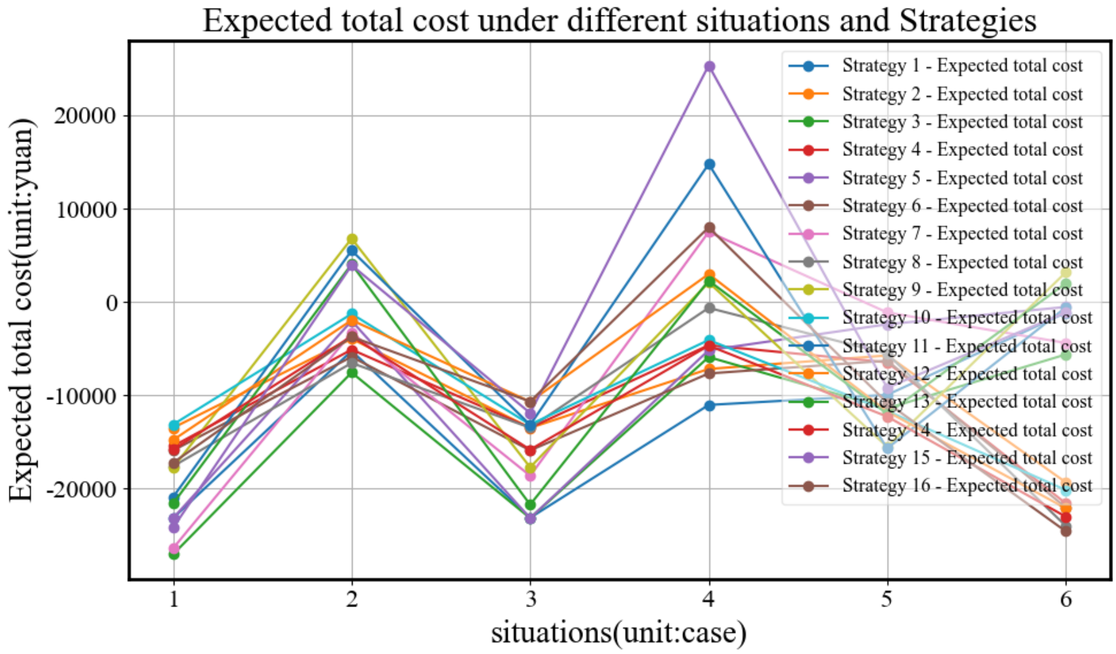


**Figure 2.** Expected total cost under different situations and Strategies

Overall, the analysis and the resulting charts provide a clear comparison between different strategies, offering a theoretical basis for enterprises to make informed decisions regarding quality control in production. These results help businesses effectively balance the costs of detection and disassembly against the potential market losses from defective products, ensuring optimal resource allocation and cost minimization in actual production environments.

## 4. Conclusions and outlooks

This paper aims to solve the problem of how to optimize the production cost through reasonable detection and disassembly strategy. For the uncertain defective rate of spare parts and finished products, we built a mathematical model based on sampling inspection, and combined with the sample size calculation formula and Monte Carlo simulation to quantify the impact of different strategies on the total cost. Firstly, the minimum sample size at 95% and 90% confidence levels is determined through the sample size calculation formula, and then the sampling process under different defective rate scenarios is simulated to evaluate the acceptance probability of enterprises. The results show that enterprises can accept products earlier when the defect rate is low, and should reject products when the defect rate is high. In addition, we determine the optimal strategy combination under different production conditions to effectively reduce the total cost. The research results have wide application value in manufacturing, quality control and other fields, and can help enterprises make optimal decisions in uncertain environment.

Although this study provides an effective tool for production decision-making, there is still room for improvement. The model is based on historical data and assumed probability distribution, ignoring complex factors such as supplier quality fluctuations. Therefore, dynamic modeling can be introduced in the future to enhance the adaptability of the model. Moreover, the accuracy of Monte Carlo simulation depends on the number of simulations, which may lead to errors due to the limitation of computational resources. In the future, more efficient algorithms or machine learning methods can be introduced to improve accuracy. The research focuses on the cost optimization of detection and disassembly. In the future, supply chain management and other factors should be considered for a more comprehensive cost analysis.

## References

[1] Fang X, Luo Q, Zhou B, et al. Research progress of automated visual surface defect detection for industrial metal planar materials[J]. Sensors, 2020, 20(18): 5136.

[2] Julious S A. Sample sizes for clinical trials[M]. chapman and hall/CRC, 2023.

[3] Luo C, Keshtegar B, Zhu S P, et al. Hybrid enhanced Monte Carlo simulation coupled with advanced machine learning approach for accurate and efficient structural reliability analysis[J]. Computer Methods in Applied Mechanics and Engineering, 2022, 388: 114218.

[4] Cao P, Shen X, Duan G, et al. Quality-integrated diagnostic platform for aerospace complex product assembly processes[J]. Computers & Industrial Engineering, 2024, 189: 109796.

[5] Harris R S. Inspection models for automotive parts remanufacture[J]. 2020.

[6] Rivera-Gómez H, Gharbi A, Kenné J P, et al. Joint optimization of production and maintenance strategies considering a dynamic sampling strategy for a deteriorating system[J]. Computers & industrial engineering, 2020, 140: 106273.

[7] Menzies T, Milton Z, Turhan B, et al. Defect prediction from static code features: current results, limitations, new approaches[J]. Automated Software Engineering, 2010, 17: 375-407.

[8] Stehman S V, Foody G M. Key issues in rigorous accuracy assessment of land cover products[J]. Remote Sensing of Environment, 2019, 231: 111199.

[9] Wang L, **ao T, Liu S, et al. Quantification of model uncertainty and variability for landslide displacement prediction based on Monte Carlo simulation[J]. Gondwana Research, 2023, 123: 27-40.

[10] Pellissetti M F, Schuëller G I. Scalable uncertainty and reliability analysis by integration of advanced Monte Carlo simulation and generic finite element solvers[J]. Computers & Structures, 2009, 87(13-14): 930-947.